第四章 朴素贝叶斯法

4.1 基本介绍

朴素贝叶斯（naive Bayes）法是基于贝叶斯定理与特征条件独立假设的分类方法【注意：朴素贝叶斯法与贝叶斯估计是不同的概念】，其是一种典型的生成学习方法。生成方法由训练数据学习联合概率分布$P(X,Y)$，然后求得后验概率分布$P(Y|X)$。具体来说，利用训练数据学习$P(X|Y)$和$P(Y)$的估计，得到联合概率分布：

$$P(X,Y)=P(Y)P(X|Y)$$

概率估计的方法可以是极大似然估计或贝叶斯估计。

4.2 基本假设

朴素贝叶斯法的基本假设是条件独立性，即:

$$\begin{aligned} P(X=x|Y=c_k) &= P(X^{(1)}=x^{(1)},\ldots,X^{(n)}=x^{(n)}|Y=c_{k)} \\ &= \prod_{i=1}^{n} P(X^{(j)}=x^{(j)}|Y=c_{k)}\} \end{aligned}$$

这是一个比较强的假设。由于这一假设，模型包含的条件概率的数量大为减少，朴素贝叶斯法的学习与预测大为简化。因而朴素贝叶斯法高效，且易于实现。其缺点是分类的性能不一定很高。【这里的更加朴素的理解是A被分类成正类和负类的两个事件是独立的】。如果假设朴素贝叶斯法中的假设输入变量都是条件是不独立的，那么此时，模型就变成了贝叶斯网络。

4.3 基本方法

朴素贝叶斯法分类时，对于给定的输入x，通过学习到的模型计算后验概率分布$P(Y=c_k|X=x)$，将**后验概率最大的类作为x的类**的输出。后验概率计算根据贝叶斯定理进行：

$$P(Y=c_k|X=x)=\frac{P(X=x|Y=c_k)P(Y=c_k)}{\sum_k P(X=x|Y=c_k)P(Y=x_k)}$$

将4.2式中假设带入到4.3中即得以下表达式：

$$P(Y=c_k|X=x)=\frac{P(Y=c_k)\prod_j P(X^{(j)}=x^{(j)}|}{\sum\limits_k P(Y=x_k)\prod_j P(X^{(j)}=x^{(j}}$$

这就是朴素贝叶斯法分类的的基本公式，朴素贝叶斯分类器可表示为

$$y=f(x)=\underset{c_k}{argmax}\frac{P(Y=c_k)\prod\limits_j P(X^{(j)}=x^{(j)}|Y=c_k)}{\sum\limits_k P(Y=c_k)\prod\limits_j P(X^{(j)}=x^{(j)}|Y=c_j)}$$

事实上，上面表达式中分母对所有c_k都是相同的，所以有

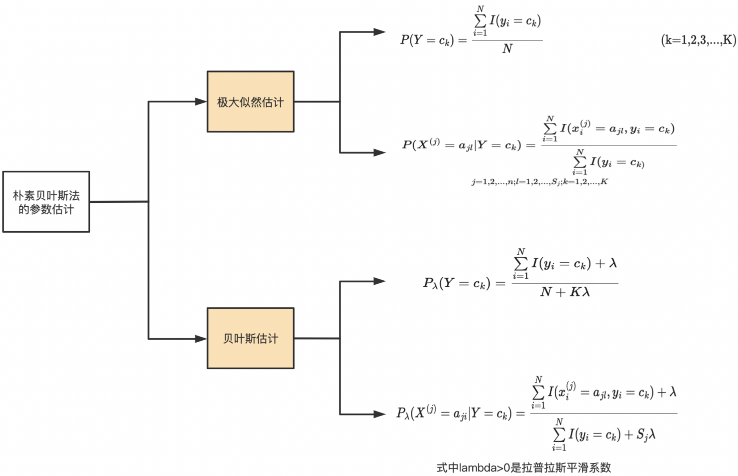
$$y=\underset{c_k}{argmax}P(Y=c_k)\prod_j P(X^{(j)}=x^{(j)}|Y=c_k) \quad (⭐)$$

> 同时，可证明后验概率最大化的含义就等价于风险最小化

4.4 朴素贝叶斯法的参数估计

通过上面⭐式，可以知道需要分别求$P(Y=c_k)$以及$P(X^{(j)}=x^{(j)}|Y=c_k)$。这里主要有两种估计方法：极大似

然估计和贝叶斯估计。



4.5 案列

对于上面公式的理解可能会有一点苦难，在下面的例子中分别用这两种方法来具体演示：根据下面的训练数据学习一个朴素贝叶斯分类器并确定$x=(2,S)^T$的类标记y。表中$X^{(1)}$和$X^{(2)}$为特征，Y为目标。

表 4.1 训练数据

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$X^{(1)}$	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3
$X^{(2)}$	S	M	M	S	S	S	M	M	L	L	L	M	M	L	L
Y	-1	-1	1	1	-1	-1	-1	1	1	1	1	1	1	1	-1

4.5.1 使用极大似然估计

由上表易计算下列概率：

$P(Y=1)=\frac{9}{15}$，$P(Y=-1)=\frac{6}{15}$

$P(X^{(1)}=1|Y=1)=\frac{2}{9}$，$P(X^{(1)}=2|Y=1)=\frac{3}{9}$，$P(X^{(1)}=3|Y$

$P(X^{(2)}=S|Y=1)=\frac{1}{9}$，$P(X^{(2)}=M|Y=1)=\frac{4}{9}$，$P(X^{(2)}=L$

$P(X^{(1)}=1|Y=-1)=\frac{3}{6}$， $P(X^{(1)}=2|Y=-1)=\frac{2}{6}$， $P(X^{(1)}=$
$P(X^{(2)}=S|Y=-1)=\frac{3}{6}$， $P(X^{(2)}=M|Y=-1)=\frac{2}{6}$， $P(X^{(2)}$
对于给定的$x=(2,S)^T$计算：
$P(Y=1)P(X^{(1)}=2|Y=1)P(X^{(2)}=S|Y=1)=\frac{9}{15}\bullet\frac{3}{9}\bullet\frac{1}{9}=$
$P(Y=-1)P(X^{(1)=2}|Y=-1)P(X^{(2)}=S|Y=-1)=\frac{6}{15}\bullet\frac{2}{6}\bullet\frac{3}{6}$
显然根据上面式子的计算，后者更大，根据朴素贝叶斯法是将实例分到后验概率最大的类中的原理，故$y=-1$

4.5.2 使用贝叶斯估计

同理，在使用贝叶斯估计下有以下表达式：
$P(Y=1)=\frac{10}{17}$， $P(Y=-1)=\frac{7}{17}$
$P(X^{(1)}=2|Y=1)=\frac{4}{12}$， $P(X^{(1)}=2|Y=-1)=\frac{3}{9}$
$P(X^{(2)}=S|Y=1)=\frac{2}{12}$， $P(X^{(2)}=S|Y=-1)=\frac{4}{9}$ 对于给定的$x=(2,S)^T$计算：
$P(Y=1)P(X^{(1)}=2|Y=1)P(X^{(2)}=S|Y=1)=\frac{10}{17}\bullet\frac{4}{12}\bullet\frac{2}{12}=$
$P(Y=-1)P(X^{(1)}=2|Y=-1)P(X^{(2)}=S|Y=-1)=\frac{7}{17}\bullet\frac{3}{9}\bullet$
显然根据上面式子的计算，后者更大，根据朴素贝叶斯法是将实例分到后验概率最大的类中的原理，故$y=-1$

4.6 sklearn API

[朴素贝叶斯]